PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Improving site selection in clinical studies: a standardized, objective, multistep method and first experience results
AUTHORS	Hurtado-Chong, Anahi; Joeris, Alexander; Hess, Denise; Blauth, Michael

VERSION 1 - REVIEW

REVIEWER	Diane Whitham University of Nottingham, United Kingdom
REVIEW RETURNED	30-Nov-2016

GENERAL COMMENTS	The second sentence in the results section of the abstract is methodology and not results I would suggest moving this to the methods section. I would like to see the results presented in the body of the paper summarised here in the abstract. In the methods section of the paper point 3 I would like to have seen the detail of the selection criteria used, this was not clear to me as a reviewer. However in the results section the selection criteria is defined in more detail " selecting for the following criteria: daily visits of geriatrician....." This should be moved to the method section. In note from the recruitment graph that the study should have completed or nearly completed recruitment. Is there a more up to date recruitment graph showing that the initial trend identified continued? In the discussion you indicate that this methodology has been used in 3 other studies, does the author have any information to support these preliminary findings?

REVIEWER	Giuseppe Ambrosio Division of Cardiology University of Perugia School of Medicine Perugia, Italy
REVIEW RETURNED	21-Dec-2016

GENERAL COMMENTS	This study aims at assessing means to improve selection of sites to run clinical trials. A method based on use of a network, definition of objective criteria, and systematic screening process, was used through an open call using a newly-developed multistep approach. Out of 267 interested sites, 12 sites were initially selected. The steps included: an open call through a network, use of selection questionnaires tailored to the study, evaluation of responses using

	objective criteria, and scripted telephone interviews. The number of candidate sites was step-wise reduced, leaving what authors considered the most promising candidate sites. Authors conclude that the results of this experience with a standardized and objective method of site selection are encouraging, in spite of some problems with contracting. COMMENTS: Delay in effectively and efficiently run clinical studies, particulalry in recruiting patients, is a common problem, which result in increased duration and costs of trials. Also, poor site selection can result in poor data quality. Optimizing site selection is therefore an interesting approaqch to improving study quality and completion. In this respect, this paper investigates a novel and relevant issue. The findings presented are interesting, and point to a potentially more efficacious way of selecting sites. I would recommend authors to consider the following points: 1. Control Group I understand the nature of this investigation, and therefore comparison against a proper control Group cannot be performed. Yet, it would be nice if authors could speculate how this new approach compare with what had been previously achieved in this field by neans of "classical" site selection processes. 2. "Weight" of criteria Many criteria were used to gauge the quality of a gven site. It would be important to know: a) how these criteria were chosen; b) whether if a different "weight" was attributed to each. 3.Accuracy of responses Clearly, most of initial screening relies on self-declaration by potentially interested sites, which would then make site pass on to the next screening step. Did you perform a post-hoc check of accuracy of such? 4. Correlation between "score" and actual perfomance We are told that many criteria were used to identify prospective sites. It would seem reasonable that not all of them were present at the same time and with the same degree in all selected sites. It would be nice to show such a distribution across sites, and to verify whether such an heterogeneity actually predicts level of performance. 5. Role of high-experience sites I'd suggest to introduce also a sobering note re considering high-experience sites as the best. It is common experience that, just because of such a high reputation, they end up being involved in multiple trials simutaneously. Even though this may not technically be considered as "competing trials", it ends up using human and technical resources for a new additional trial. 6. Other factors Please, discuss that site selection, although quite important, is not the only hurdle toward effectively performing large trials. Other concomitant factors (e.g., length of approval, contracting, etc...) play an equal or more important role. See: Gehring M, et al. Factors influencing clinical trial site selection in Europe: the Survey of Attitudes towards Trial sites in Europe (the SAT-EU Study). British Medical Journal Open 15;3(11):e002957.

	doi: 10.1136/bmjopen-2013-002957, 2013

REVIEWER	Dr Emma Ogburn University of Oxford UK
REVIEW RETURNED	22-Dec-2016

GENERAL COMMENTS	This site selection method appears a lengthy process - would be good to include the timeframe of when the initial email call was sent to site opened and possibly first patient recruited and how much time sites need to invest before they are accepted/declined, only the questionnaire has a time. the sites circumstance such as competing trials could have changed if its a lengthy process. also, the initial call via email, not many opened the email, why? Africa is listed as one of the countries - is internet connection an issue, especially a stable connection to complete the questionnaire. does this system bias against research naïve sites or smaller sites who initially aren't as knowing initially as larger, research active sites but can sometimes end up being the best recruiter?

REVIEWER	Carl Fichtenbaum University of Cincinnati College of Medicine United States
REVIEW RETURNED	03-Jan-2017

GENERAL COMMENTS	This is an interesting article detailing methods for choosing sites for clinical research. The crux of this study lies within the methods to "filter out sites". The authors do not specifically indicate what criteria they used to filter a site out after the first round in the methods. Instead they detail these criteria within the results. It is not clear if a site was missing one criteria, were they excluded? There is no data presented in tabular form showing differences between sites selected and not selected for each round. Also, were there additional criteria, For example, is there a minimum number of patients with a particular problem required? Is there a minimum number of surgeries required? The authors need to provide more details on what specific criteria they used to eliminate those that completed a questionnaire and then completed a phone interview. The reader needs to know what criteria they used to go from 267 potential sites to 12 sites. Second, the authors indicate that 2 sites were dropped (16%) and were replaced. And the study is ongoing with accrual not yet complete. This raises a question as to whether the methods used have any impact? How do we know that this method works any better than any other methods used? Also, it looks like there is substantial variability in the accrual. Four of 12 sites have accrual under 10 subjects per group. Is this acceptable? Ideally, if the authors are correct in their methods, their accrual would be completed within their proposed time interval or sooner. And the quality of data would be very good. There is no information presented on the quality of the data by the sites. We have no metrics on data submitted, data errors, lost to follow up, queries, protocol violations, etc. So while the idea seems interesting it leaves the reader with too many unanswered questions. We need data on

	quality of the information from the sites and we need much more data presented on process and sites excluded. And since this is not randomized, how will we know that 12 other sites not selected by the process would not have performed just as well? It is encouraging to see the accrual curve appears to be on target. It would help to know when the study was proposed to start and when it actually started. On page 5 the authors state (lines 89-92): Unlike pharmaceutical trials, in which administration of a drug might be relatively easy to train, clinical studies involving surgeries demand not only patients willing to participate but also the doctors' expertise on the particular technique(s) or devices under study, and in occasions advanced infrastructure. This statement is a bit demeaning to investigators that conduct drug investigation trials. I would suggest the authors modify this statement to indicate that skills are required for surgical studies without denigrating those who conduct pharmaceutical trials. Lines 109-113 are not strictly detailing methods and more a description of rationale. This probably belongs more in an introduction or discussion section and should be removed from the methods.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Diane Whitham
1. The second sentence in the results section of the abstract is methodology and not results I would suggest moving this to the methods section. I would like to see the results presented in the body of the paper summarised here in the abstract.

This is a methodological paper and as such the aim is to describe the process used. We have used the GFC study just as an illustrative example

2. In the methods section of the paper point 3 I would like to have seen the detail of the selection criteria used, this was not clear to me as a reviewer. However in the results section the selection criteria is defined in more detail " selecting for the following criteria: daily visits of geriatrician....." This should be moved to the method section.

Within the methods section, the specific criteria used to select GFC or UCC centers were not detailed but were anyhow referenced as Supplementary 2, section 3 (pg 6, line 145). If this paper was presenting the results of the GFC study or even the protocol, the selection criteria would have been definitively detailed in the methods because it is an important part of the methodology. However, the present manuscript is not about the GFC study, but rather a methodological paper which aims to introduce a novel approach to site selection and in which the GFC study is used as an illustrative example of how the method was developed and applied. The main message is that the criteria used for selection have to be tailored to the study and there must be a differentiation between mandatory requirements and desirable ones. This had been already mentioned in the discussion (pg 12, line 293-296): "It is absolutely mandatory to adapt the template to include the assessment of specific aspects of the study, identifying strict requirements from desirable ones. Strict aspects will become the first focal point for selection at the beginning of the assessment helping to reduce the number of applications rather quickly". In addition, clarifying phrases have been added to the Methods section (pg 6 lines 142-145).
The specific selection criteria are the result of tailoring the questions to the study design and defining

the mandatory requirements. For this reason, we still believe that the detailed selection criteria should not be specifically mentioned within the methodology section but rather in the results. Given that this methodological paper is slightly different to a research one, and to avoid misleading the reader we have changed the heading of Results for "Illustrative example of the method at work: first experience results"

3. In note from the recruitment graph that the study should have completed or nearly completed recruitment. Is there a more up to date recruitment graph showing that the initial trend identified continued?
Recruitment for this study is finished and respective figures have been updated to include the final results of the accrual
In the discussion you indicate that this methodology has been used in 3 other studies, does the author have any information to support these preliminary findings?
The studies referenced in which this method has been used are still on a very early stage and therefore cannot provide the required information at this point.

Reviewer: 2
Reviewer Name: Giuseppe Ambrosio
1. Control Group
I understand the nature of this investigation, and therefore comparison against a proper control Group cannot be performed. Yet, it would be nice if authors could speculate how this new approach compare with what had been previously achieved in this field by neans of "classical" site selection processes.

Our organization has been performing clinical research for the last 15 years. Over this time, we had studies with low recruitment in several occasions and strategies to overcome it had to be implemented during the course of these studies: extend the recruitment period beyond its originally planned time, adding new sites while closing those with poor performance or even cancelling a study before accrual completion. All of the above resulted in extended timeframes and higher costs. This explanation has already been added (pg 5 lines 98-105).
There is evidence in the literature that low accrual is rather common and one of the main reasons to discontinue a clinical trial. Kasenda et al found that nearly 10% of clinical trials were discontinued due to poor recruitment (Kasenda et al 2014). Up to 18% of cancer trials in phase II/III closed with low accrual or were recruiting less than 50% of the target after 3 years of initiation (Bennette 2016). Qureshi et al (2012) found that 78% of participating sites involved in acute stroke trials enrolled < 2% of participants and Schroen et al (2012) reported that 29% of Oncology trials in phase III closed due to poor accrual. We have included a statement about low recruitment reported in the literature (pg. 11 line 270-271). Although all the above information cannot be used as a direct comparator to our study given the differences in population, disease, study design and that the specific strategies for site selection are unknown, it highlights the deep impact that underperforming sites have on clinical studies.

2. "Weight" of criteria
Many criteria were used to gauge the quality of a given site. It would be important to know: a) how these criteria were chosen; b) whether if a different "weight" was attributed to each.

Although an interesting idea, we did not attributed weights or scores for each criteria.For this particular study, our first focal point of selection was based on sites which could qualify as GFC or UCC following the predefined criteria for each type of center. Then, we identified the countries in which both modalities co-existed and finally the information given during the telephone interview. We have included this suggestion within the discussion (pg 11 line 296-298)

3.Accuracy of responses

Clearly, most of initial screening relies on self-declaration by potentially interested sites, which would then make site pass on to the next screening step. Did you perform a post-hoc check of accuracy of such?

One first approach is to contact the potential sites and perform a telephone interview to verify the information (pg 7, line 161-163). In the particular case of GFC, the guidelines used by the hospital for their geriatric care were requested and revised (pg 7, line 166-167 and pg10 lines 172-173). Another check occurs when the site initiation visit is performed. During this visit, our personnel meet the research team, visit the hospital, and get to know its infrastructure and organization. Accuracy of declared numbers on patient number is normally not checked as it would require looking into patient files if there is no database or statistics available from the hospital. Finally, we are guided by the performance of the sites, their responsiveness to data queries and quality control performed during monitoring visit. In the discussion we had already pointed out that "Site selection visits can be a great added value to assess the site, personnel and infrastructure, although it was not performed for this study due to budget constraints (pg 12 line 298-300)". Following the reviewer's comment "accuracy of responses" has been added to this sentence.

4. Correlation between "score" and actual perfomance
We are told that many criteria were used to identify prospective sites. It would seem reasonable that not all of them were present at the same time and with the same degree in all selected sites. It would be nice to show such a distribution across sites, and to verify whether such an heterogeneity actually predicts level of performance.

As stated on question 2, we did not have a scoring system. Our best indicator of overall performance is that recruitment finished according to timelines. The strict selection criteria are present in all sites as summarized in figure 3.

5. Role of high-experience sites
I'd suggest to introduce also a sobering note re considering high-experience sites as the best. It is common experience that, just because of such a high reputation, they end up being involved in multiple trials simutaneously. Even though this may not technically be considered as "competing trials", it ends up using human and technical resources for a new additional trial.

We do believe that previous experience in clinical research and past performance provide very good information about the site. As such, we regard experience as a desirable, but it is not the most important aspect and accordingly, experience was evaluated during the last phases of the process. We agree with the reviewer that high- experience sites which might have participation in multiple trials simultaneously could hamper the resources available for new studies, and for this reason in our site selection questionnaires we always ask how many studies have been performed in the last 5 years and if there are currently other studies in the same field being performed or planned. This explanatory note has been introduced (pg 7 line 174-177)

6. Other factors
Please, discuss that site selection, although quite important, is not the only hurdle toward effectively performing large trials. Other concomitant factors (e.g., length of approval, contracting, etc...) play an equal or more important role. See:
Gehring M, et al. Factors influencing clinical trial site selection in Europe: the Survey of Attitudes towards Trial sites in Europe (the SAT-EU Study). British Medical Journal Open 15;3(11):e002957. doi: 10.1136/bmjopen-2013-002957, 2013

Length of approval and contracting problems, regulatory problems had already been mentioned in the introduction and discussion and in our view these are features to be considered within the site

selection process and not independently since these aspects even when in the same country vary widely between sites. However a clarifying sentence and the reference have been added (pg 4, lines 79-81)

Reviewer: 3

This site selection method appears a lengthy process - would be good to include the timeframe of when the initial email call was sent to site opened and possibly first patient recruited and how much time sites need to invest before they are accepted/declined, only the questionnaire has a time. the sites circumstance such as competing trials could have changed if its a lengthy process. also, the initial call via email, not many opened the email, why? Africa is listed as one of the countries - is internet connection an issue, especially a stable connection to complete the questionnaire. does this system bias against research naïve sites or smaller sites who initially aren't as knowing initially as larger, research active sites but can sometimes end up being the best recruiter?

From the first call to the first site opened to recruitment, the full protocol was developed along with supplementary documentation such as the patient information sheet, informed consent, patient diaries, etc. The criteria for GFC were reviewed and approved by the Foundation´s research commission and once the short list of preselected sites was ready it also underwent review and approval. Next translations into the appropriate languages were performed prior to submission to the ethics committees. For all the above, the timeframe from the initial call to the opening of the first site for recruitment is affected by several factors in our case and therefore is not a reliable indicator. Although it might seem that this lengthy process might play against sites who have competing trials, given that it is an iterative process, it is more likely to detect changes of situation/interest during the successive rounds as opposed to just doing the application and waiting for a long time for the results. This point has been addressed in the discussion (pg. 12 line 283-286)
The initial call was sent as an e-mail blast to all members of the network. Our system is able to detect how many people opened this e-mail and that is why we can provide that number. We can do nothing else but speculate on the reasons why many did not open this e-mail blast: lack of interest in research by practicing orthopaedic surgeons, lack of interest in the topic as it might just call the attention of a few and not to the whole community, e-mail went into spam folder, e-mail is not checked by the owner or is inactive, lack of time to see or open e-mails. Unfortunately there is no way to verify this information.
As any on-line platform, REDCap, relies on internet connection. We run all our studies in REDCap, therefore internet connection is a requirement to qualify as a study site. However, as it can be seen in their partner list, REDCap has several projects and users from Africa.
When choosing a site, several aspects are weighted in and what deemed important or mandatory will depend on the particularities of the study. In our case, we consider a plus if a site shows a good track record however it is not the most important aspect but it could tick the balance towards them compared to another site with the same characteristics without experience.

Reviewer: 4
Reviewer Name: Carl Fichtenbaum
1. The crux of this study lies within the methods to "filter out sites". The authors do not specifically indicate what criteria they used to filter a site out after the first round in the methods. Instead they detail these criteria within the results. It is not clear if a site was missing one criteria, were they excluded? There is no data presented in tabular form showing differences between sites selected and not selected for each round. Also, were there additional criteria, For example, is there a minimum number of patients with a particular problem required? Is there a minimum number of surgeries required? The authors need to provide more details on what specific criteria they used to eliminate those that completed a questionnaire and then completed a phone interview. The reader needs to know what criteria they used to go from 267 potential sites to 12 sites.

This is a methodological paper and for this reason some information which would normally appear in the methodology section of a research article are placed within the results. Our method requires to define upfront which requirements are mandatory. In our case, the specific criteria for the selection of either a GFC center or a UCC are a result of applying such method by defining the mandatory requirements. For this reason, we believe that this information belongs in the results. We have included a figure and a table to illustrate how each specific criterion filtered out sites and how many sites per country remained after applying these strict criteria considered as mandatory. Following this, other reasons to eliminate sites were already described in the text (pg 10, lines 220-245). We have changed the heading of Results for "Illustrative example of the method at work: first experience results"

Second, the authors indicate that 2 sites were dropped (16%) and were replaced. And the study is ongoing with accrual not yet complete. This raises a question as to whether the methods used have any impact? How do we know that this method works any better than any other methods used? Also, it looks like there is substantial variability in the accrual. Four of 12 sites have accrual under 10 subjects per group. Is this acceptable? Ideally, if the authors are correct in their methods, their accrual would be completed within their proposed time interval or sooner. And the quality of data would be very good. There is no information presented on the quality of the data by the sites. We have no metrics on data submitted, data errors, lost to follow up, queries, protocol violations, etc. So while the idea seems interesting it leaves the reader with too many unanswered questions. We need data on quality of the information from the sites and we need much more data presented on process and sites excluded. And since this is not randomized, how will we know that 12 other sites not selected by the process would not have performed just as well? It is encouraging to see the accrual curve appears to be on target. It would help to know when the study was proposed to start and when it actually started.

We believe our method works because compared to our own previous experience and the information available in the literature regarding low accrual this first experience gave good results. Our recruitment phase finished already and indeed our accrual was completed within the expected timeframe. However, we extended the enrollment a bit further to allow each site to have at least 20 patients as explained in pg 11, line 258-263. We have updated the figures with the final numbers. Certainly this is our first experience with this method and has shown good results, but the robustness of this method can only be determined with further experience.

3. On page 5 the authors state (lines 89-92):
Unlike pharmaceutical trials, in which administration of a drug might be relatively easy to train, clinical studies involving surgeries demand not only patients willing to participate but also the doctors' expertise on the particular technique(s) or devices under study, and in occasions advanced infrastructure. This statement is a bit demeaning to investigators that conduct drug investigation trials. I would suggest the authors modify this statement to indicate that skills are required for surgical studies without denigrating those who conduct pharmaceutical trials.

4. The first part of the sentence has been removed. It now reads as follows (pg 5 line 90-93): "Clinical research in surgically-related topics faces several challenges. It requires not only patients willing to participate but also the doctors' skills and expertise on the particular technique(s) or devices under study, and in occasions advanced infrastructure"

VERSION 2 – REVIEW

REVIEWER	Diane Whitham University of Nottingham
REVIEW RETURNED	22-Mar-2017

GENERAL COMMENTS	This paper appears to be a descriptive piece rather than a results driven methodology paper, I believe this should be made clearer to the reader. In the results (p.11 line 264) n the author states "....all sites have shown good compliance and high quality data". I would like to see the statement supported with some data, e.g completeness of data, return rates, number of queries, loss to follow-up by site etc. even if the study is ongoing it would have been good to see some preliminary information on these metrics. I would still like to understand the criteria applied to the site selection. I understand that the trial will have had specific inclusion criteria but did the team work to pre-specified criteria when selecting sites? For example number of site staff to ongoing trial ratios, number of potential eligible participant, etc...

REVIEWER	Giuseppe Ambrosio University of Perugia School of Medicine Perugia, Italy
REVIEW RETURNED	06-Mar-2017

GENERAL COMMENTS	Authors have satisfactorily addressed my comments

VERSION 2 – AUTHOR RESPONSE

Response to reviewer 1:

1. This paper appears to be a descriptive piece rather than a results driven methodology paper, I believe this should be made clearer to the reader.

The last phrase of the introduction has been changed as follows (pg5, lines 105-109): "The purpose of the present article is to describe the methodology developed to standardize our site selection process for multicenter clinical studies. The GFC study is used as an example to illustrate how the method works and the results obtained with this first experience."

2. In the results (p.11 line 264) the author states "....all sites have shown good compliance and high quality data". I would like to see the statement supported with some data, e.g completeness of data, return rates, number of queries, loss to follow-up by site etc. even if the study is ongoing it would have been good to see some preliminary information on these metrics.

A summary of number of patients enrolled, eligibility failures, drop outs and monitoring follow up rate has been included (pg. 11 line 254-263): "Of the 282 patients enrolled, no eligibility failures have been detected and so far only 36 patients have dropped out, mainly due to death (24/36). The overall monitoring follow up rate of the study is defined as the number of visits done divided by the sum of visits due plus visits done (excluding dropouts) and is used as an indicator of quality. For the first study visit at 12 weeks, the overall monitoring follow up rate of the study is 84% ranging from 50% to

97% depending on the study site. Sites with low follow up rates might still improve if the reason for the currently missing visits is because the information in the study database has not been updated yet. The GFC study has a large and complex database that collects up to 1,484 variables for each patient but notably, our overall response rate to queries is currently 96%"

The other reasons for drop out were withdrawal of consent (7/36), and lost to follow up (5/36). The current monitoring FU rate at 1 year is 87% although this is still very preliminary as for many patients the visit is not due yet.

3. I would still like to understand the criteria applied to the site selection. I understand that the trial will have had specific inclusion criteria but did the team work to pre-specified criteria when selecting sites? For example number of site staff to ongoing trial ratios, number of potential eligible participant, etc...

For this study, the first and most important limiting factor was to find GFC and having a matching site for comparison within the same country. We found that one of the most challenging aspects in this study was to identify GFC and UCC. Therefore this became our first focal point as it has been described in the methods section (pg 8 line 198): the first step to make the selection was to identify GFC and UCC centers. Next we selected for centers that had both modalities of care within the same country. These two aspects turned to be the most challenging and as described in the paper it already decreased the number of candidate sites in a considerable way. As explained in the corresponding section, the list continued to be further reduced after considering participation in competing clinical trials, double checking the available guidelines, lack of disposition, discrepancies with the information provided previously, high fees, difficult regulatory conditions, etc.
In contrast to other studies in which this could normally be the first selection criteria, in the particular case of the GFC study, the number of patients was not a limiting factor given our broad inclusion criteria: patients older than 70 years with an osteoporotic hip fracture treated surgically. The range of declared patients treated per year varied between 40-600 so we considered that those numbers would be sufficient to reach the recruitment target of 23-25 patients within 16 months, and as such this parameter ultimately did not play an effective role in selecting sites. Likewise, research staff available did not play a role for selection in this study. Nevertheless it is important to point out that even if it didn't turn out to be a crucial factor for the GFC study, patient population normally is the first focal point for selection in other studies. The bottomline is that it is very important to identify the strict requirements from the desirable ones so the strict ones become the first focal point for selection as we illustrate in this example.

VERSION 3 – REVIEW

REVIEWER	Diane Whitham University of Nottinghamn
REVIEW RETURNED	19-Apr-2017

GENERAL COMMENTS	Thank you for responding to my queries.